IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 February 2024, accepted 26 February 2024, date of publication 4 March 2024, date of current version 12 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3373437

RESEARCH ARTICLE

Robust Subject-Wise Dictionary Learning for fMRI

MUHAMMAD USMAN KHALID AND BADER M. ALBAHLAL
College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University, Riyadh 11564, Saudi Arabia
Corresponding author: Muhammad Usman Khalid (mukhalid@imamu.edu.sa)

ABSTRACT This paper presents a robust subject-wise sequential dictionary learning (swsDL) algorithm named rswsDL for functional magnetic resonance imaging (fMRI) data where the negative impact of dimensionality reduction in the form of information loss and sensitivity to anomalous observations due to the assumption of Gaussian prior has been resolved by rotating the reduced dimensions to optimal direction and replacing the quadratic loss in the data fidelity term with α-divergence based loss function to counter the outliers, respectively. While dimensionality optimization guaranteed robustness to model order by maximizing signal intensity and smoothness, the robust loss function guaranteed decomposition stability against deviations from the Gaussian noise model. The proposed algorithm was derived by deploying spatial and temporal bases from the computationally fast sparse spatiotemporal blind source separation (ssBSS) method and solving a sequence of rank-1 matrix decomposition problems, where the l_1/l_0-norm penalty/constraint promoted sparsity, and the estimation of sparse representation matrices was accomplished using a block coordinate descent approach. This strategy allowed the utilization of multi-subject fMRI data to enhance the subject-wise source separation in a robust manner. It, therefore, can be considered a promising alternative to the state-of-the-art robust consistent adaptive sequential dictionary learning (rACSD) algorithm. The rswsDL and existing robust dictionary learning based source separation algorithms were applied to synthetic and experimental fMRI datasets to validate its performance. The rswsDL algorithm manifested a 16.7% increase in the mean correlation value over the rACSD algorithm.

INDEX TERMS Robust subject-wise analysis, fMRI, sparse representation, group-level analysis, dictionary learning, representation matrices, signal intensity maximization, α-divergence.

I. INTRODUCTION

Detecting the activated brain region during cognition and functionally connected brain regions during rest is challenging due to the fMRI signal's poor signal-to-noise ratio (SNR) [1]. Functional MRI data is traditionally analyzed by deploying the computationally expensive seed voxel-based cross-correlation analysis [2], or statistical parametric mapping (SPM) toolbox [3], which employs a hypothesis-driven methodology by modelling the hemodynamic response function (HRF) for the univariate general linear model. In contrast, data-driven approaches like blind source separation (BSS) have provided many effective results for fMRI analysis in recent years [4], [5], [6], [7], [8], [9] when experimental responses diverge from the expected response and cannot be accurately replicated due to differences among subjects and brain regions [10]. Particularly, resting state functional connectivity analysis and task-based activation detection have benefited greatly from using multivariate data matrix decomposition techniques [11], [12], [13], [14], [15], [16], [17].

Principal component analysis (PCA) is used as a preprocessing step for dimensionality reduction in mixing matrix-based BSS algorithms like ICA [18], canonical correlation analysis (CCA) [19], and partial least square (PLS) [20]. While keeping too many PCs may lead to unreliable separation results by BSS algorithms such as ICA that may generate phantom source signals consisting of spikes and

The associate editor coordinating the review of this manuscript and approving it for publication was Easter Selvan Suviseshamuthu.

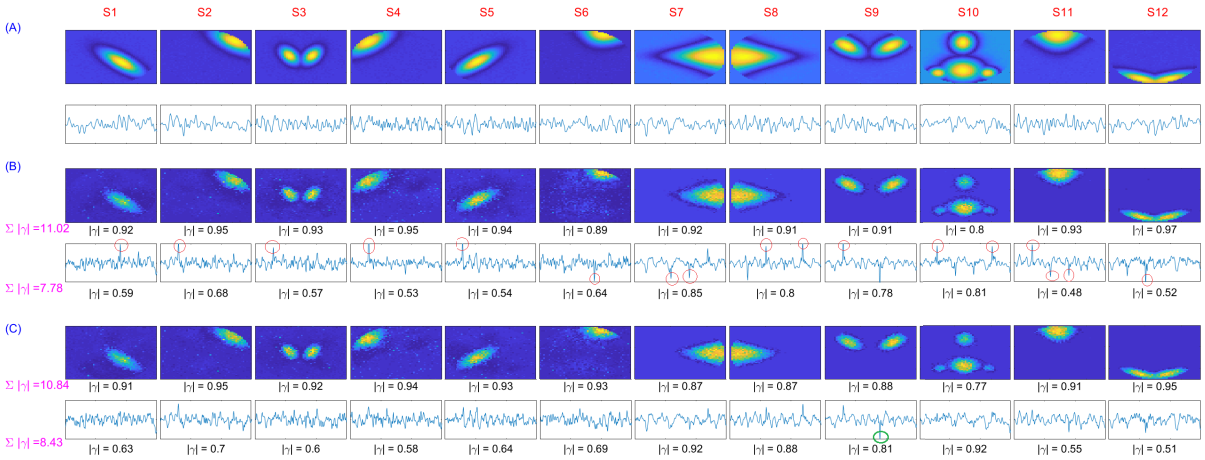

FIGURE 1. A) Generated spatial and temporal sources, B) spatial and temporal sources recovered by rswsDL$_2$ (rswsDL with l_2 data fidelity term), and C) spatial and temporal sources recovered by rswsDL$_1$ (rswsDL with α-divergence based data fidelity term). The absolute temporal and spatial correlation values (γ) are shown against each source, along with the sum of correlation values on the left. The red circles on the temporal sources obtained from rswsDL$_2$ represent the outliers that were successfully suppressed by rswsDL$_1$, but were not attenuated by rswsDL$_2$. In contrast, green circles on the temporal sources recovered by rswsDL$_1$ suggest those outliers that both methods failed to suppress.

bumps [21], strict reduction leads to information loss and may seriously compromise the performance of the subsequent source separation algorithm [2]. Thus, any mismatch between the retained dimension and the actual model order can significantly impact the source retrieval results. Recently, this has been addressed using the signal intensity maximization (SIM) technique [22], [23] to increase the resilience of BSS models to the order number. In particular, more information is retained using PCA, while SIM helps the ensuing BSS algorithm steer clear of the overfitting issue. In order to reduce the sensitivity of ssBSS to the model order, the proposed algorithm incorporates dimensionality optimization via basis expansion.

The traditional DL methods assume the Gaussian prior for noise that leads to a quadratic loss function as a maximum likelihood estimate (MLE). Due to its susceptibility to outliers, l_2 norm data fitting term is usually replaced with least absolution deviation, Huber loss, or truncated l_1 norm error according to the assumptions made about the outlier statistics [24]. By deploying these tools from robust statistics [25], the effect of outliers can be mitigated for dictionary learning. For instance, the online dictionary learning algorithm is incorporated with robust function through l_1 data fidelity term, where an update for each dictionary sample is obtained using weighted correlation matrices and conjugate gradient method for various computer vision tasks [26], a capped l_1 norm-based loss function is used for robust DL to iteratively solve the re-weighted least square problem for face recognition [27], a robust DL algorithm handles outliers based on each atom's update frequency through confidence-weighted sparse code and fine-tuned dictionary [28].

Instead of developing an outlier identification framework for the fMRI data [29], [30], [31], [32], our goal is to fully utilize the statistical information of the outliers and design an algorithm that would diminish their effects on the data representations in the form of a dictionary [33], [34], [35]. Therefore, we define an outlier problem as the time points in the fMRI series containing high noise/artifacts/outliers levels due to magnetic field instabilities, head movement, physiological effects, and fMRI data complex preprocessing steps. We aim to design an outlier removal scheme that has an automatic suppression ability while estimating the unknowns to down-weigh all those datapoints that are far from the bulk of the data. One way to approach this is by assuming noise as non-Gaussian and deploying class of divergences other than Kullback-Leibler such as α-divergence [34] that will provide a loss function that is adapted to address robustness against anomalous observations. For the synthetic fMRI data in this paper, we attempted to generate these outliers in the form of a Laplacian noise.

A. INTRODUCTORY EXAMPLE

Outliers are generally acknowledged to occur in fMRI data and are not entirely removed even after the pre-processing stages [29], [30], [31], [32]. Thus, we characterize an outlier problem as the presence of time points in the fMRI time series that exhibit elevated amounts of noise, artifacts, and outliers. These anomalies might arise from magnetic field instabilities, head movement, physiological effects, and the many preprocessing procedures for handling fMRI data.

To verify how data fidelity term's modification address the negative impact of outliers, we conducted an experiment in which anomalies were introduced in the synthetic dataset

by means of a Laplacian noise. The proposed rswsDL algorithm with data fidelity term derived from α-divergence (rswsDL$_1$) was successfully able to learn dictionary atoms from synthetic datasets that were contaminated by strong Laplacian noise, and so was rswsDL with l_2 data fidelity term (rswsDL$_2$), however rswsDL$_1$ performed better. To highlight the difference between these two competing algorithms, we tried to recover twelve different source signals and maps from the datasets generated using Laplacian noise and some outliers that were manually added at time points 60, 120, 180, 240.

For an unbiased comparison, all tuning parameters, dictionary size, and algorithm iterations were set the same for both algorithms. One is encouraged to read this paper's section IV-D for more details on generating these datasets, selecting values for the parameters, and extracting the sources. From Fig, 1C, one can conclude that rswsDL$_1$ is successful in eliminating 95% of the outliers and in the process of doing so it came out with the higher overall temporal correlation strength than rswsDL$_2$.

II. RELATED WORK

In this paper, the robustness of the proposed algorithm was enhanced by reducing the information loss using the SIM technique and attenuating the effect of anomalous observations using the α-divergence based loss function. While a SIM-based dimensional optimization allowed retaining more information from the data without any concern about overfitting, α-divergence provided robustness against noise by down-weighing the anomalies in the data. Moreover, motivated by the recent approach that utilized autoencoder to project multi-subject fMRI data to a shared space before subject-specific decoders reconfigured it [36], the robust dictionaries were trained using a mixing model where the mixing matrices were adapted to the subject-wise training data as discussed in [37]. This approach enhanced the statistical significance of the recovered sources by combining the spatiotemporal diversities from multiple subjects.

The proposed model for robust loss function differs from the existing models [34] as it is characterized by i) data utilization in reduced dimension space for computational efficiency, ii) subject-wise dictionary that embodies multi-subject spatiotemporal diversity, and iii) dictionary with discrete cosine transform (DCT) bases and spatiotemporal mixing matrices. A two-level strategy is used for the proposed dictionary training, where the first level trains a base dictionary and the base sparse code using an efficient ssBSS method with the SIM approach. The second level estimates mixing matrices using the sequential approach with α-divergence directed rank-1 decomposition, soft thresholding, autocorrelation maximization, and paired minimization.

The rest of the paper is divided into four sections. Section III describes the proposed preliminaries, the proposed model, and the proposed algorithm, followed by section IV, presenting the synthetic and experimental studies consisting of comparisons and performance metrics. The paper ends with section V, which provides concluding remarks. This is followed by a section on acknowledgement.

III. MATERIALS AND METHODS

While retaining the simplicity of l_2 minimization [38], the robust ACSD algorithm [34] has shown superior source recovery performance in the presence of non-Gaussian noise. However, its limited applicability to subject-wise analysis produced spatiotemporal dynamics that lacked strength due to its inability to exploit statistical dependencies among groups of subjects. In this paper, hemodynamic variations from multiple subjects are utilized computationally efficiently to increase the accuracy and strength of subject-wise robust analysis.

A. ROBUST ssBSS

Recently, a novel optimization model for source separation of single-subject fMRI signals was presented [39]. This led to the development of the sparsity-based source separation method (ssBSS), which uses alternating least squares and soft thresholding to solve for the unknown model variables via Neumann's alternating projection lemma. Using multi-subject data, the proposed rswsDL algorithm in this paper builds a spatial and temporal base dictionary by means of the subject-wise ssBSS method due to its computational efficiency.

It is noteworthy at this point that the rswsDL model defined by equation (16) gains from the shared inherent structures across brains while performing subject-wise source extraction. In order to get those inherent structures, we could either rely on predefined paradigms that do not possess experimental variations across brains and their regions or use DCT/spline bases that also could not provide any experimental variations. The issue lies in the fact that both of them are static priors and are incapable of adapting to variations. One way variations across brains and regions and shared structures across brains could have been considered was to deploy two rounds of dictionary learning or data matrices as a base matrix. Both strategies would make the resultant algorithm computationally expensive and impractical. Hence, utilizing ssBSS's temporal and spatial components to construct the base spatial and temporal matrices turned out to be computationally very convenient due to its numerical efficiency, as discussed and shown in [39].

The ssBSS method seeks to decompose the m-th subject fMRI whole-brain dataset Y_m into the temporal source matrix $\mathrm{T}_m \in \mathbb{R}^{N\times P}$ and the spatial source matrix $\mathrm{S}_m \in \mathbb{R}^{P\times V}$ as follows

$$\begin{aligned} &\min_{\mathrm{C}_m,\mathrm{S}_m} \left\|\mathrm{Y}_m - \mathrm{T}_p\mathrm{C}_m\mathrm{S}_m\right\|_F^2 + \lambda_1 \left\|\mathrm{S}_m\right\|_1, \\ &\text{sub.to.} \quad \left\|\mathrm{T}_p \mathrm{c}_{m,p}\right\|_2 = 1, \quad \left\|\mathrm{c}_{m,p}\right\|_0 \le \zeta_1 \end{aligned} \tag{1}$$

where $m = \{1,\ldots,M\}$, M is the number of subjects, $\mathrm{T}_m = \mathrm{T}_p\mathrm{C}_m$, the DCT bases are stored in $\mathrm{T}_p \in \mathbb{R}^{N\times K_p}$ to realize the smoothness of the BOLD signal, C_m is the

sparse representation matrix and $c_{m,p}$ is its p-th column. Sparsity is promoted via $\|.\|_0$ and $\|.\|_1$ norm that counts the number of non-zero elements and the sum of absolute values, respectively, and λ_1/ζ_1 are the sparsity hyperparameters. An efficient solution can be produced by disintegrating equation (1) into a pair of spatial and temporal sources as

$$\min_{C_m,Q_m} \left\|\Omega_{m,t} - T_p C_m Q_m\right\|_F^2,$$
$$\text{sub.to.} \quad \left\|T_p c_{m,p}\right\|_2 = 1, \quad \left\|c_{m,p}\right\|_0 \le \zeta_1 \tag{2}$$
$$\min_{S_m,Z_m} \left\|\Omega_{m,s} - Z_m S_m\right\|_F^2 + \lambda_1 \left\|S_m\right\|_1 \tag{3}$$

where $Q_m \in \mathbb{R}^{P\times K}$ and $Z_m \in \mathbb{R}^{K\times P}$ are the mixing matrices, and $\Omega_{m,t} \in \mathbb{R}^{N\times K}$ and $\Omega_{m,s} \in \mathbb{R}^{K\times V}$ obtained after dimensional optimization consist of the temporal and spatial features in the reduced dimension, respectively. They are obtained as $\Omega_{m,t} = X_{m,t}W$ and $\Omega_{m,s} = WX_{m,s}$ using the demixing matrix $W \in \mathbb{R}^{K\times K}$, which consists of the eigenvectors of $U + U^\top$, and U can be computed as

$$U = gX_{m,t}^\top \tilde{X}_{m,t} + (1-g)\Gamma^2/\gamma_1 \tag{4}$$

In order to assign weight to the autocorrelation terms and signal existence terms, a parameter g that exists in the interval of $(0, 1)$ was used. Here, matrix Γ contains the eigenvalues of $Y_m Y_m^\top$, γ_1 is the most significant eigenvalue, and $\tilde{X}_{m,t}$ is one time-sample delayed version of the following problem

$$\min_{\tilde{C}_m} \left\|X_{m,t} - T_p \tilde{C}_m\right\|_F^2,$$
$$\text{sub.to.} \quad \left\|T_p \tilde{c}_{m,k}\right\|_2 = 1, \quad \left\|\tilde{c}_{m,k}\right\|_0 \le \zeta_2 \tag{5}$$

and $X_{m,t}$ and $X_{m,s}$ are obtained using singular value decomposition (SVD) i.e. $Y = \Sigma\Gamma\Delta^\top$ as

$$X_{m,t} = Y\Delta\Gamma^\top, \quad X_{m,s} = \Sigma^\top Y$$

In summary, a high reduction number was chosen when performing SVD, indicating that a larger number of dimensions were preserved through the process of dimensionality reduction. This decision was made in order to maintain a greater amount of valuable information. Then, the SIM technique was used to mitigate the issue of overfitting effectively, which achieves this by aligning the reduced dimensions towards the optimal orientation, enabling it to maximize the presence of interesting signals. Consequently, it can differentiate between genuine existing components and non-existent artifacts, particularly when an incorrect number of principal components are chosen. This is equivalent to the estimation of demixing matrix using equation (4).

The ssBSS method was applied individually to all participating subjects in the study to learn spatial and temporal components that very precisely resemble the underlying ground truth sources as explained in [39]. The extracted temporal components are concatenated horizontally to build up a base dictionary matrix, and extracted spatial components are concatenated vertically to build up a base sparse code matrix. Both multi-subject base matrices will be utilized by the rswsDL method to perform subject-wise source separation and they are given as

$$D_q = [T_1, T_2, \ldots, T_M]$$
$$X_q = [S_1^\top, S_2^\top, \ldots, S_M^\top]^\top \tag{6}$$

B. ROBUST SUBJECT-WISE swsDL

1) α-DIVERGENCE

Conventional dictionary learning (DL) algorithms assume a Gaussian noise prior, leading to the maximum likelihood estimate producing a squared l_2 term. This loss term is vulnerable to outliers and does not provide any protection against them. From probabilistic point of view, we also know that maximizing the likelihood is equivalent to minimizing the statistical distance between the true probability distribution and the estimated one, which is the definition of Kullback-Leibler divergence (KLD). Instead of KLD, we can consider a different type of divergence, such as α-divergence [34]. This lets us have a loss function that is better at handling anomalous observations than the quadratic loss function.

As described in [34], consider an independent identically distributed data sample $r_1, \ldots, r_V$, where i-th $r_i \in \mathbb{R}^N$ is sampled from a probability distribution $\mathcal{G}(r, \hat{\omega})$ with $\omega \in \Omega \subset \mathbb{R}^\kappa$. Let $\mathcal{G}(r, \omega), \omega \in \Omega$ corresponds to the family of approximating probability distributions parameterized by ω. For the corresponding probability density functions $g(r, \omega)$ and $g(r, \hat{\omega})$, the α-divergence [40], which is a measure of similarity between two probability distributions offers protection against anomalous observations unlike the Kullback-Leibler divergence, and is given as

$$D_\alpha\Big(g(r, \hat{\omega}) \parallel g(r, \omega)\Big) = \frac{1}{\alpha(\alpha-1)}\left(\int g(r, \hat{\omega})^\alpha g(r, \omega)^{1-\alpha} dr - 1\right) \tag{7}$$

This is the generalization of KLD indexed by the parameter α, which can be observed by deriving D_α in terms of the α-logarithmic function as

$$\log_\alpha(x) = \frac{x^{(1-\alpha)} - 1}{1-\alpha} \tag{8}$$

and minimizing the α-divergence with respect to ω is the same as maximizing the following

$$\hat{\omega} = \arg\max_\omega \frac{1}{n}\sum_{i=1}^{n} \log_\alpha\{g(r_i, \omega)\} \tag{9}$$

Its detailed derivation is described in [34]. Now, considering the probabilistic swsDL model $y_i = DABx_i + \eta_i$, where $\eta_i \in \mathbb{R}^N$ is a zero mean Gaussian white residual vector with variance σ^2. Considering its j-th entry η_i^j as r_i^j then

$$g(\eta_i^j, \omega) = g(y_i^j - d^j ABx_i) \propto \exp\left(-\frac{(y_i^j - d^j ABx_i)^2}{2\sigma^2}\right) = \exp\left(-\frac{\eta_i^{j2}}{2\sigma^2}\right) \tag{10}$$

TABLE 1. Algorithm for solving the minimization problem (16).

Given:
Training set $(\mathrm{Y}_m \in \mathbb{R}^{N\times V}, m = 1, ..., M)$
Base dictionary and sparse code $\mathrm{D}_q/\mathrm{X}_q$ obtained using eq. (6)
Tuning parameters $(\lambda_2, \zeta_2, \zeta_3, \alpha, \mu)$, Number of iterations (L)
Constants $(K, \epsilon = 0.01, l = 0)$

1. **Initialize**
   $\mathrm{A}_m \leftarrow \mathrm{I}_{MP\times K}, \mathrm{B}_m \leftarrow 0, \mathrm{D}_m \leftarrow \mathrm{D}_q\mathrm{A}_m, \mathrm{X}_m \leftarrow \mathrm{B}_m\mathrm{X}_q, L \leftarrow 15$
   $\mathrm{W}_i \leftarrow \mathrm{I}_{N\times N} \quad \forall \quad i = 1, \ldots, V$

2. **for** $l \leftarrow 1$ to L
   if $l \neq 1$
   Compute W_i as $\exp\left(-\alpha(\mathrm{y}_{m,i} - \mathrm{D}_m\mathrm{x}_{m,i})^2/2\right)$
   end

3. **for** $k \leftarrow 1$ to K

4. **Compute**
   $\mathrm{a}_{m,k} \leftarrow 0, \mathrm{b}_m^k \leftarrow 0, \mathrm{x}_m^k \leftarrow 0, \mathrm{d}_t \leftarrow 0, t \leftarrow 0$
   $f_{m,i}^1 \leftarrow 0, \quad f_{m,i}^n \leftarrow d_{m,i}^{n-1}, \quad i = 1, .., K, \quad n = 2, .., N$
   $\mathrm{Z}_m \leftarrow \frac{1}{\mu+2}(\mathrm{F}_m^\top\mathrm{F}_m)^{-1}\mathrm{F}_m^\top\mathrm{D}_m\mathrm{X}_m$
   $\mathrm{E}_m^k \leftarrow \mathrm{Y}_m - \mathrm{D}_m\mathrm{X}_m + \mathrm{F}_m\mathrm{Z}_m$

5. **Repeat while** $\|\mathrm{d}_t - \mathrm{d}_{m.k}\|_2 > \epsilon$ & $t \leq 3$
   $\mathrm{d}_t \leftarrow \mathrm{d}_{m,k}$
   $t \leftarrow t + 1$

6. **Update** b_m^k
   $\mathrm{x}_m^k \leftarrow \mathrm{d}_{m,k}^\top\mathrm{E}_{m,k}$
   $\beta = \mathrm{a}_{m,k}^\top\mathrm{D}_q^\top\mathrm{W}_i\mathrm{e}_{m,i}^k/\mathrm{a}_{m,k}^\top\mathrm{D}_q^\top\mathrm{W}_i\mathrm{D}_q\mathrm{a}_{m,k}$
   $\lambda_{2,i}^k = \lambda_2/\mathrm{a}_{m,k}^\top\mathrm{D}_q^\top\mathrm{W}_i\mathrm{D}_q\mathrm{a}_{m,k}$
   $\mathrm{x}_{m,i}^k = \mathrm{sgn}(\beta) \circ (|\beta| - \lambda_{2,i}^k/2)_+, \quad i = 1, \ldots, V$
   Compute ϑ using thresholding correlation algorithm in [39] and ζ_3
   $\mathrm{b}_{m,\vartheta}^k \leftarrow \mathrm{x}_m^k\mathrm{X}_q^{\vartheta\top}\left(\mathrm{X}_q^\vartheta\mathrm{X}_q^{\vartheta\top}\right)^{-1}$
   $\mathrm{x}_m^k \leftarrow \mathrm{b}_m^k\mathrm{X}_q$

7. **Update** $\mathrm{a}_{m,k}$
   $\mathrm{d}_{m,k} = \left(\sum_{i=1}^{V}\mathrm{W}_i(\mathrm{b}_m^k\mathrm{x}_{q,i})^2\right)^{-1}\left(\sum_{i=1}^{V}\mathrm{W}_i\mathrm{e}_{m,i}^k\mathrm{b}_m^k\mathrm{x}_{q,i}\right)$
   Compute ϑ using thresholding correlation algorithm in [39] and ζ_2
   $\mathrm{a}_{m,k}^\vartheta \leftarrow (\mathrm{D}_{q,\vartheta}^\top\mathrm{D}_{q,\vartheta})^{-1}\mathrm{D}_{q,\vartheta}^\top\mathrm{d}_{m,k}$
   $\mathrm{a}_{m,k} \leftarrow \mathrm{a}_{m,k}/\left\|\mathrm{D}_q\mathrm{a}_{m,k}\right\|_2$
   $\mathrm{d}_{m,k} \leftarrow \mathrm{D}_q\mathrm{a}_{m,k}$

8. **end**
9. **end**
10. **end**

Output: D_m and X_m

where d^j is the j-th row of D and x_i is the i-th column of X and ω represents either D or x_i based on the sparse dictionary learning stage. Using the definition of $\log_\alpha$ from equation (8) and replacing the probability density function in (10) by equation (9) we get

$$\hat{\omega} = \arg\max_{\omega} \frac{1}{N}\sum_{i=1}^{N}\frac{\exp\left(-\frac{\tau}{2\sigma^2}\eta_i^2\right) - 1}{\tau}$$
$$= \arg\min_{\omega} \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tau}\left(1 - \exp\left(-\frac{\tau}{2}\left(\frac{\eta_i}{\sigma}\right)^2\right)\right)$$

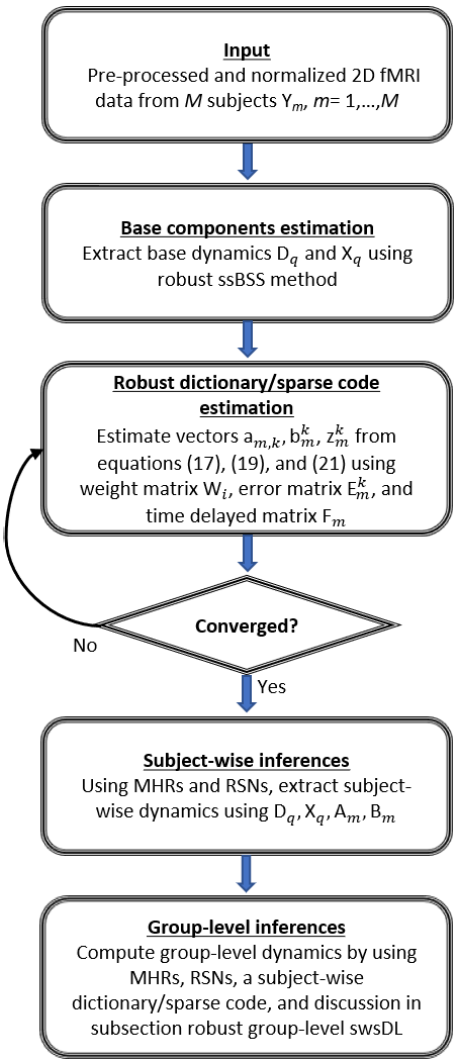

FIGURE 2. The block diagram of the proposed rswsDL algorithm showing all the steps from preprocessing to subject-wise and group-level inferences.

$$= \arg\min_{\omega} \frac{1}{N}\sum_{i=1}^{N}\ell_\tau(\nu_i) \tag{11}$$

where $\nu_i = \eta_i/\sigma$ and $\ell_\tau(\nu_i)$ is given as

$$\ell_\tau(\nu_i) = \frac{1}{\tau}\left(1 - \exp\left(-\frac{\tau\nu_i^2}{2}\right)\right) \tag{12}$$

When $\tau \to 0$, $\ell_\tau(\nu_i) \to \ell_0(\nu_i) = \nu_i^2/2$ and the well-known quadratic loss related to the Frobenius norm is recovered, which is quite susceptible to the presence of outliers in the data. The $\tau > 0$ refers to a weighted estimator that tends to down-weight errors that deviate significantly from the nominal density, eliminating outliers and promoting inference stability during the learning process. The weights' structure automatically down-weighs anomalous data or outliers that are remote from the majority of the data and have no bearing on the final estimate, making the DL resilient against outliers.

More specifically, consider the function $\exp\Big(-(e^k_{m,i} - d_{m,k}x^k_{m,i})\Big)$ from equation (16), which are the entries of the square weight matrix W_i of size $N \times N$, and this appears everywhere (equation (18) and (20) because new estimates are solved using a weighted least squares). Assume if entries of $e^k_{m,i}$ are far from $d_{m,k}x^k_{m,i}$ then a low value is assigned to the weight matrix compared to the entries that are close to $d_{m,k}x^k_{m,i}$. Thus, by down-weighing all the large numbers, all data points far from the bulk of the data will automatically be down-weighted and will have no impact on the estimates, making rswsDL robust to outliers.

It must be noted that instead of computing W_i as $\exp\Big(-\alpha(e^k_{m,i} - d_{m,k}x^k_{m,i})^2/2\Big)$, $(i = 1, \ldots, V)$ for each atom/sparse code update of the proposed algorithm, we computed a single W_i as $\exp\Big(-\alpha(y_{m,i} - D_m x_{m,i})^2/2\Big)$, $(i = 1, \ldots, V)$ for all K updates during each algorithm iteration. This resulted in some computational gain by a factor of K, and only little variations in performance regarding algorithm convergence and final results were observed.

2) swsDL

The proposed algorithm utilizes the multi-subject fMRI data by means of base matrices D_q/X_q that store spatial and temporal components from ssBSS method and representation matrices A_m/B_m that mix and demix those components. This mixing and demixing allows the combining of common variations and the separating of unique ones. Combining common variations from multiple subjects for a single subject data decomposition gives statistical strength to subject-wise dictionary learning. This is because it allows subject-wise data to be decomposed into multi-subject dictionary and sparse code.

The subject-wise dictionary learning algorithm, in general, as discussed in [37], endeavours to solve the following optimization problem based on the Frobenius norm by considering all observed signals given as

$$\begin{aligned} &\min_{A_m, B_m} \left\| Y_m - D_q A_m B_m X_q \right\|_F^2 + \lambda \left\| b^k_m X_q \right\|_1, \\ &\text{sub.to.} \quad \left\| D_q a_{m,k} \right\|_2 = 1 \end{aligned} \tag{13}$$

where it was assumed that fMRI dataset $Y_m \in \mathbb{R}^{N\times V}$ from m-th subject can be decomposed into a subject-wise dictionary $D_m \in \mathbb{R}^{N\times K}$ and subject-wise coefficient matrix $X_m \in \mathbb{R}^{K\times V}$. These subject-wise matrices are estimated using multi-subject smooth dictionary $D_q \in \mathbb{R}^{N\times MP}$, the multi-subject sparse code $X_q \in \mathbb{R}^{MP\times V}$ and their respective representation matrices $A_m \in \mathbb{R}^{MP\times K}$ and $B_m \in \mathbb{R}^{K\times MP}$ that encourage overfitting. This leads to $D_m = D_q A_m$ and $X_m = B_m X_q$. However, instead of directly decomposing the observed data matrix or error matrix of all signals, which may lead to the K-SVD algorithm [41] that has comparatively poor recovery and convergence rate, the rank-1 minimization that takes error matrix of all signals into account and promotes adaptive sparse penalty term for a fairer assignment of penalty to each entry in $b^k_m x_{q,j}$ was considered. Moreover, the representation matrices A_m and B_m were updated column-wise and row-wise, respectively, in a sparse manner to discourage overfitting given as

$$\begin{aligned} \{a_{m,k}, b^k_m\} = \arg \min_{a_{m,k}, b^k_m} &\left\| E^k_m - D_q a_{m,k} b^k_m X_q \right\|_F^2 \\ &+ \sum_{j=1}^{V} \lambda^k_{2,j} |b^k_m x_{q,j}|, \quad \text{sub.to.} \quad \left\| a_{m,k} \right\|_0 \le \zeta_2, \\ &\left\| b^k_m \right\|_0 \le \zeta_3, \quad \left\| D_q a_{m,k} \right\|_2 = 1 \end{aligned} \tag{14}$$

where each entry of $b^k_m X_q$ has a data-driven regularization parameter $\lambda^k_{2,j}$ assigned to it, sparsity for representation matrices was promoted through l_0 norm with ζ_2/ζ_3 their sparsity controlling parameters, and the error matrix for updating the representation matrices for k-th dictionary atom/sparse code was equated as

$$E^k_m = Y_m - \sum_{i=1, i\neq k}^{K} d_{m,i} x^i_m \tag{15}$$

3) PROPOSED MODEL

The proposed rswsDL algorithm is a promising alternative to rACSD because it utilizes diversities across brains. There are some variations to the ACSD method, such as ShSSDL [42] and sgBACES [43] that do allow capturing cross-subject and subject-specific variations, but the main drawback of these methods is their applicability being limited to task-related data only and inability to handle outliers. The proposed method can handle both, although we do not provide resting-state data analysis in this paper to avoid increasing the paper length.

The loss function derived in equation (12), developed in the previous subsection, was employed in the fidelity term in place of the Frobenuis norm in equation (14) to produce the proposed robust DL model as

$$\begin{aligned} \{a_{m,k}, b^k_m\} = \arg \min_{a_{m,k}, b^k_m} &\sum_{i=1}^{V} \ell_\tau \left(\frac{e^k_{m,i} - D_q a_{m,k} b^k_m x_{q,i}}{\sigma} \right) \\ &+ \sum_{i=1}^{V} \lambda^k_{2,i} |b^k_m x_{q,i}|, \quad \text{sub.to.} \quad \left\| a_{m,k} \right\|_0 \le \zeta_2, \\ &\left\| b^k_m \right\|_0 \le \zeta_3, \quad \left\| D_q a_{m,k} \right\|_2 = 1 \end{aligned}$$

where $e^k_{m,i}$ is the i-th column of E^k_m. To account for the lag-1 error minimization for the dictionary to train atoms with maximum autocorrelation, two variables $F_m \in \mathbb{R}^{N\times K}$ and $Z_m \in \mathbb{R}^{K\times V}$ were introduced. Each column of F_m was one time-sample delayed version of atoms, and the respective sparse code estimate was stored in the m-th coefficient matrix Z_m. After taking this into account, and assuming that the transformed variates are uncorrelated $Z^\top_m F^\top_m F_m Z_m = I$, the above equation can be modified to obtain the proposed model

for the robust swsDL given as

$$\{\mathrm{a}_{m,k}, \mathrm{b}_m^k, \mathrm{z}_m^k\} = \arg \min_{\mathrm{a}_{m,k}, \mathrm{b}_m^k, \mathrm{z}_m^k} \sum_{i=1}^{V} \ell_\tau \left(\frac{\mathrm{e}_{m,i}^k - \mathrm{D}_q \mathrm{a}_{m,k} \mathrm{b}_m^k \mathrm{x}_{q,i}}{\sigma} \right) + \left\| \mathrm{D}_q \mathrm{a}_{m,k} \mathrm{b}_m^k \mathrm{X}_q - \mathrm{F}_m \mathrm{Z}_m \right\|_F^2 + \sum_{i=1}^{V} \lambda_{2,i}^k |\mathrm{b}_m^k \mathrm{x}_{q,i}|, \quad \text{sub.to.} \quad \left\| \mathrm{a}_{m,k} \right\|_0 \leq \zeta_2, \quad \left\| \mathrm{b}_m^k \right\|_0 \leq \zeta_3, \quad \left\| \mathrm{D}_q \mathrm{a}_{m,k} \right\|_2 = 1, \quad \mathrm{Z}_m^\top \mathrm{F}_m^\top \mathrm{F}_m \mathrm{Z}_m = \mathrm{I} \tag{16}$$

We suggest using iterative optimization to solve the objective in (16), which attempts to estimate one rank-1 matrix out of K at a time. For this particular case, update for $\mathrm{a}_{m,k}$, b_m^k, and z_m^k was obtained by first finding the updates for $\mathrm{d}_{m,k}$ and x_m^k. These two variables are solved as a pair by performing K penalized rank-1 approximations of (16).

4) DICTIONARY REPRESENTATION MATRIX UPDATE

With b_m^k and z_m^k fixed, an update for a_m^k was obtained by solving the Lagrangian of (16) with respect to a_m^k as

$$\mathrm{a}_{m,k} = (\mathrm{D}_q^\top \mathrm{D}_q)^{-1} \mathrm{D}_q^\top \mathrm{d}_{m,k}$$

By considering the l_0 constraint on a_m^k above equation as a constrained problem becomes

$$\mathrm{a}_{m,k} = \arg \min_{\mathrm{a}_{m,k}} \left\| \mathrm{d}_{m,k} - \mathrm{D}_q \mathrm{a}_{m,k} \right\|_2^2 \quad \text{sub.to.} \quad \left\| \mathrm{a}_{m,k} \right\|_0 \leq \zeta_2 \tag{17}$$

where the unknown $\mathrm{d}_{m,k}$ was solved by considering

$$\mathrm{d}_{m,k} = \min_{\mathrm{d}_{m,k}} \sum_{i=1}^{V} \left(\frac{1}{\tau} - \frac{1}{\tau} \exp \left(- \frac{\tau (\mathrm{e}_{m,i}^k - \mathrm{d}_{m,k} \mathrm{b}_m^k \mathrm{x}_{q,i})^2}{2\sigma^2} \right) \right)$$

and differentiating it w.r.t $\mathrm{d}_{m,k}$ to obtain

$$\sum_{i=1}^{V} \frac{1}{\sigma} \left(\mathrm{d}_{m,k} (\mathrm{b}_m^k \mathrm{x}_{q,i})^2 - \mathrm{e}_{m,i}^k \mathrm{b}_m^k \mathrm{x}_{q,i} \right) \exp \left(- \frac{\tau (\mathrm{e}_{m,i}^k - \mathrm{d}_k \mathrm{b}_m^k \mathrm{x}_{q,i})^2}{2\sigma^2} \right) = 0$$

Thus, the solution for $\mathrm{d}_{m,k}$ was obtained as

$$\mathrm{d}_{m,k} = \left(\sum_{i=1}^{V} \mathrm{W}_i (\mathrm{b}_m^k \mathrm{x}_{q,i})^2 \right)^{-1} \left(\sum_{i=1}^{V} \mathrm{W}_i \mathrm{e}_{m,i}^k \mathrm{b}_m^k \mathrm{x}_{q,i} \right) \tag{18}$$

This was followed by l_2 normalization of $\mathrm{d}_{m,k}$. Here $\mathrm{W}_i \in \mathbb{R}^{N \times N}$ is a square weight matrix with values $\exp(-\tau(\mathrm{e}_{m,i}^k - \mathrm{d}_{m,k} \mathrm{x}_{m,i}^k)^2 / 2\sigma^2)$ on the diagonal, and for simplicity σ was assumed as 1.

5) SPARSE CODE REPRESENTATION MATRIX UPDATE

With a_m^k and z_m^k fixed, and solving the Lagrangian of (16) with respect to b_m^k, its update was obtained as

$$\mathrm{b}_m^k = \mathrm{x}_m^k \mathrm{X}_q^\top \left(\mathrm{X}_q \mathrm{X}_q^\top \right)^{-1}$$

Above equation can be rewritten as a constrained problem when the l_0 constraint on b_m^k was considered

$$\mathrm{b}_m^k = \arg \min_{\mathrm{b}_m^k} \left\| \mathrm{x}_m^k - \mathrm{b}_m^k \mathrm{X}_q \right\|_2^2 \quad \text{sub.to.} \quad \left\| \mathrm{b}_m^k \right\|_0 \leq \zeta_3 \tag{19}$$

where the i-th entry of the unknown x_m^k was solved by considering

$$\mathrm{x}_{m,i}^k = \min_{\mathrm{x}_{m,i}^k} \sum_{i=1}^{V} \left(\frac{1}{\tau} - \frac{1}{\tau} \exp \left(- \frac{\tau (\mathrm{e}_{m,i}^k - \mathrm{D}_q \mathrm{a}_{m,k} \mathrm{x}_{m,i}^k)^2}{2\sigma^2} \right) \right)$$

and taking its derivative w.r.t $\mathrm{x}_{m,i}^k$ and setting it to zero to obtain

$$\frac{1}{\sigma} \mathrm{a}_{m,k}^\top \mathrm{D}_q^\top \mathrm{W}_i (\mathrm{e}_{m,i}^k - \mathrm{D}_q \mathrm{a}_{m,k} \mathrm{x}_{m,i}^k) + \lambda_{2,i}^k \mathrm{sgn}(\mathrm{x}_{m,i}^k)) = 0$$

where W_i is a square weight matrix as described in the previous subsection and the solution for $\mathrm{x}_{m,i}^k$ was obtained as

$$\mathrm{x}_{m,i}^k = \mathrm{sgn}(\beta) \circ (|\beta| - \lambda_{2,i}^k / 2)_+ \tag{20}$$

where $\beta = \mathrm{a}_{m,k}^\top \mathrm{D}_q^\top \mathrm{W}_i \mathrm{e}_{m,i}^k / \mathrm{a}_{m,k}^\top \mathrm{D}_q^\top \mathrm{W}_i \mathrm{D}_q \mathrm{a}_{m,k}$ and $\lambda_{2,i}^k = \lambda_2 / \mathrm{a}_{m,k}^\top \mathrm{D}_q^\top \mathrm{W}_i \mathrm{D}_q \mathrm{a}_{m,k}$.

6) COEFFICIENT MATRIX UPDATE

By fixing a_m^k and b_m^k, a block update of Z from (16) that uses the whole dictionary and coefficient matrices rather than just the most recent modified atom/sparse code was obtained by solving the Lagrangian of

$$\mathrm{Z}_m = \arg \min_{\mathrm{Z}_m} \left\| \mathrm{D}_q \mathrm{A}_m \mathrm{B}_m \mathrm{X}_q - \mathrm{F}_m \mathrm{Z}_m \right\|_F^2, \quad \text{sub.to.} \quad \mathrm{Z}_m^\top \mathrm{F}_m^\top \mathrm{F}_m \mathrm{Z}_m = \mathrm{I}$$

and the closed form solution was obtained as

$$\mathrm{Z}_m = \frac{1}{\mu + 2} \left((\mathrm{F}_m^\top \mathrm{F}_m)^{-1} \mathrm{F}_m^\top \mathrm{D}_q \mathrm{A}_m \mathrm{B}_m \mathrm{X}_q \right). \tag{21}$$

where μ is the Lagrangian multiplier. A variation of the iterative re-weighted least squares method for rank-1 matrix approximation can be observed in the resulting dictionary/sparse code update.

C. ROBUST GROUP-LEVEL swsDL

The suggested model was applied to multiple subjects; hence, group-level analysis was conducted in addition to the subject-wise data decomposition. For this purpose, canonical HRF from the SPM toolbox and task stimuli were used to create modelled HRFs (MHRs) for task-related analysis, and resting-state network templates (RSNs) were obtained from [44] for resting-sate analysis. While equation (6) was used to prepare base dictionary and sparse

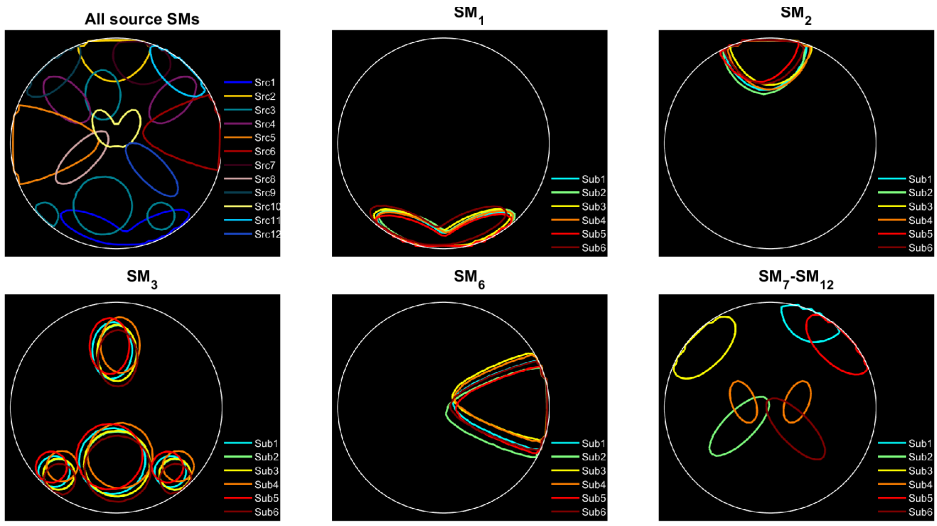

FIGURE 3. All spatial sources of the first subject in the first subfigure, whereas the location, shape, and variability of some spatial sources across subjects and unique spatial sources are shown in the rest of the subfigures.

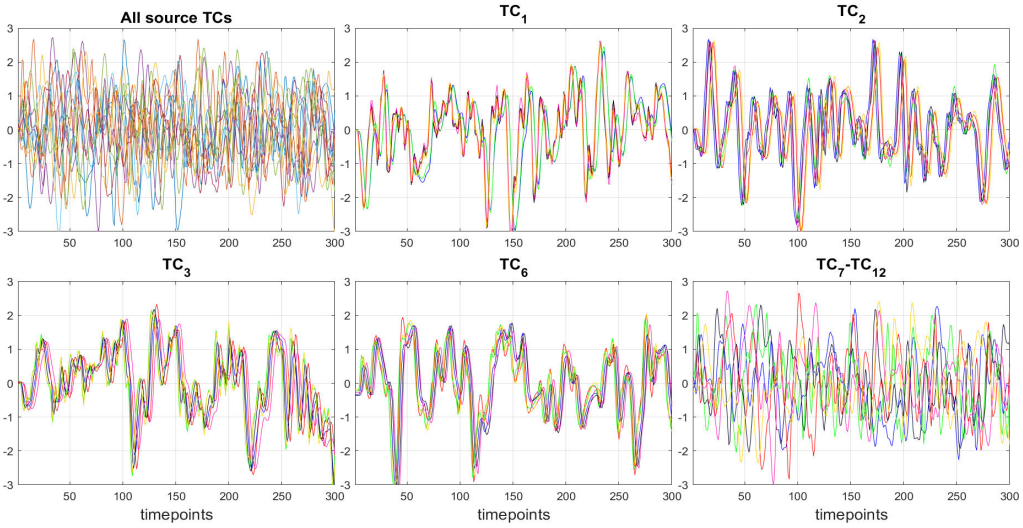

FIGURE 4. Respective temporal sources of Fig. 3 where the first subfigure contains the timecourses of the first subject and some of the common sources having temporal variability with unique temporal sources shown in rest of the subfigures.

code, equation (17), (18), (19), (20), and (21) were used to obtain subject-wise dictionary and sparse code. This was followed by group analysis, where firstly, the task-related group components were estimated by concatenating dictionary atoms/sparse codes according to the neuroscience knowledge as $\mathrm{D}_r = [\mathrm{d}_{1,j_1(r)}, \mathrm{d}_{2,j_2(r)}, \ldots, \mathrm{d}_{M,j_M(r)}]$ and $\mathrm{X}_r = \left[\mathrm{x}_1^{j_1(r)^\top}, \mathrm{x}_2^{j_2(r)^\top}, \ldots, \mathrm{x}_M^{j_M(r)^\top}\right]^\top$. Secondly, the rank-1 decomposition was applied on the concatenated dictionary atoms and sparse codes as $\frac{1}{M}[\mathrm{D}_r\mathrm{X}_r] = \omega_r\delta_r\gamma_r^\top$, $\mathrm{d}_{g,r} = \omega_r$, and $\mathrm{x}_g^r = \delta_r\gamma_r^\top$. Here $r = \{1, \ldots, R\}$, R represents the number of MHRs, $j_m(r)$ signifies the indices of the most correlated atom in m-th dictionary with r-th MHR, $m = \{1, \ldots, M\}$, M is the number of subjects, D_g is the group-level dictionary, and X_g is the group-level sparse code. In contrast, only sparse code was assembled for the resting-state components as $\Pi_r = \left[\mathrm{x}_1^{j_1(r)^\top}, \mathrm{x}_2^{j_1(r)^\top}, \ldots, \mathrm{x}_M^{j_M(r)^\top}\right]^\top$ and $\mathrm{x}_g^r = \frac{1}{M}\sum_{m=1}^{M}\pi_r^m$. The robust swsDL algorithm for estimating robust dictionary and sparse code is described in Table 1, and its block diagram is shown in Fig. 2.

IV. EXPERIMENTS

In this section, the proposed algorithm was assessed to determine how well it performed compared to the most advanced robust dictionary learning based data-driven algorithms already in use. In this regard, two distinct fMRI datasets, one synthetic and one experimental, were used for the data analysis. The synthetic fMRI dataset of six subjects was produced using the Simtb toolbox [45]. In contrast, the

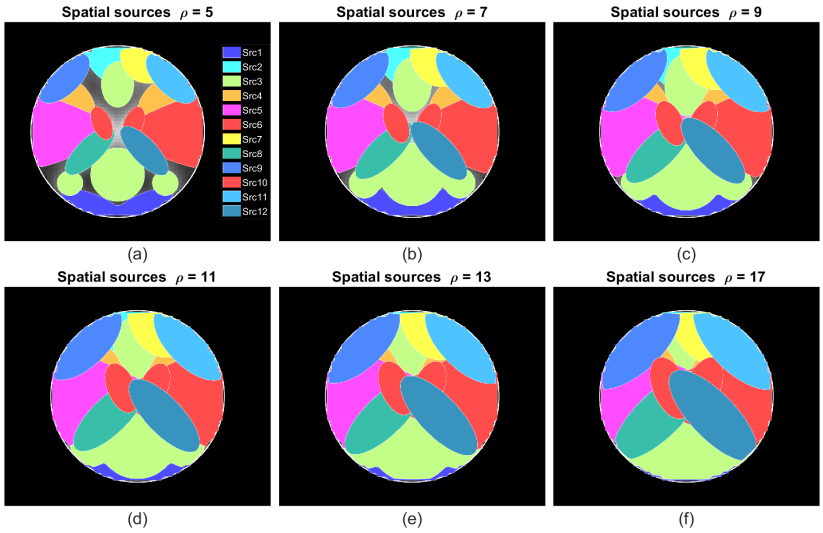

FIGURE 5. Moderate to large spatial overlaps were formed by altering the size of twelve unique activation blobs using seven different values of the spread parameter ρ, of which only 6 are shown here.

block design fMRI dataset for eight subjects was acquired from the quarter 3 release of the Human Connectome Project (HCP) [46], [47].

The comparative study consisted of the robust sparse group ICA (rsgICA) [23], [48], [49], robust group ssBSS using residual matrices (sRMSSE) [50], robust KSVD (rKSVD) [33], robust ACSD (rACSD) [34], and the proposed robust swsDL (rswsDL) algorithm. Each algorithm's performance was evaluated in terms of how well it could recover the ground truth using the aforementioned datasets.

A. SYNTHETIC DATASET GENERATION

Same as the other studies [37], [39], [50], the Simtb toolbox was used to generate a dataset of six subjects in MATLAB that mimicked the experimental fMRI data. Twelve different temporal sources, consisting of 300 time-points with a repetition time of TR = 1 sec, and twelve different spatial sources, consisting of 50 × 50 voxels each, were used to construct the fMRI dataset for six subjects. For the generation of spatial components, we used {3, 6, 8, 10, 22, 23, 26, 30, 4, 12, 5, 29} as the source IDs, which allowed mapping the synthetic brain. Among all source signals, seven spatiotemporal sources were utilized to generate the fMRI data for each individual. While the last 6 sources were distinct, and one was assigned to each subject, the first six spatiotemporal sources were common and present in all six participants with some degree of intersubject variability, as shown in more detail in [37].

Regarding common temporal sources, the HRF parameters, such as delay, response dispersion, and undershoot, were changed to induce heterogeneity among subjects. Similar to this, the Gaussian distribution parameters (mean (μ) and standard deviation (std) (σ)) were used to establish the intersubject variability for the common spatial maps. These parameters allowed for control over the activations' position, orientation, and spread. This was accomplished using the random translation ($\mu = 0, \sigma = 1$) in both the x and y directions, random rotation ($\mu = 0, \sigma = 0.9$), and random scaling ($\mu = \rho, \sigma = 0.05$) as illustrated in Fig. 3. Here, the spread parameter empowered to govern the spatial extent of the activations and produced seven distinct examples of spatial overlaps as $\rho = \{5, 7, 9, 11, 13, 15, 17\}$, the mean of the Gaussian distribution. Fig. 5a-f displays the most relevant spatial maps with moderate to large spatial dependence.

TABLE 2. Group-level extraction results in terms of mean correlation values for twelve spatiotemporal sources (ST1 -ST12) over 30 trials and different values of ρ with the highest results highlighted in bold.

Sources	rsgICA	sRMSSE	rKSVD	rACSD	rswsDL
ST1	0.737	0.704	0.735	0.750	**0.752**
ST2	0.767	**0.771**	0.755	0.729	0.758
ST3	0.801	0.772	0.823	0.800	**0.830**
ST4	0.712	0.759	0.825	0.774	**0.854**
ST5	0.827	0.847	0.867	0.844	**0.876**
ST6	0.810	0.830	0.883	0.841	**0.888**
ST7	0.477	0.695	0.598	0.661	**0.792**
ST8	0.628	0.674	0.633	0.531	**0.787**
ST9	0.585	0.710	0.648	0.612	**0.800**
ST10	0.565	0.665	0.618	0.537	**0.773**
ST11	0.583	0.683	0.623	0.691	**0.802**
ST12	0.542	0.652	0.554	0.636	**0.738**
Mean	0.670	0.730	0.713	0.701	**0.804**

The common and unique spatiotemporal sources of the first subject and five additional unique spatiotemporal sources from the remaining five subjects were used to assemble the ground truth time courses (TCs) and spatial maps (SMs) for group-level analysis, as shown in Fig. 3 and 4 as all source TCs/SMs. The remaining subfigures display some of the temporal and spatial sources that were utilized to create each of the six datasets, which was expressed as $\mathrm{Y}_m = \sum_{i=1}^{7}(\mathrm{tc}_i + \psi_i)(\mathrm{sm}^i + \phi^i)$. The Laplacian distributions $\sim \mathcal{L}(0,\ 0.9)$ and $\sim \mathcal{L}(0,\ 0.01)$ were used to create the noise-generating matrices, $\Psi \in \mathbb{R}^{300\times 7}$ and $\Phi \in \mathbb{R}^{7\times 2500}$, respectively. Then, depending

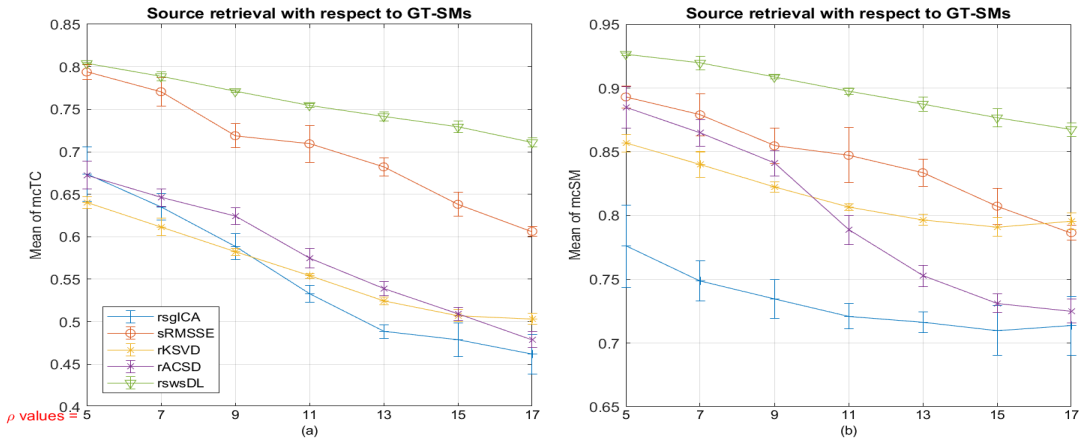

FIGURE 6. Subject-wise analysis' mean values of (a) cTC and (b) cSM computed across 7 components, 30 realizations, and 6 subjects with noise variance 0.9 and 7 values of spread parameter. Error bars represent the departure from the mean values.

on the values of ρ and trial number, the datasets $\mathrm{Y}_{m=1}^{M}$ (where $M = 6$) were composed and used by all source retrieval algorithms.

B. SYNTHETIC DATASET DICTIONARY LEARNING

The parameter values were kept the same for all algorithms to create an unbiased comparison wherever possible. The total number of components needed to be retrieved was 12 since the rsgICA was applied to the grouped data. For other algorithms, however, it was considered that the precise number of underlying sources for real fMRI data is unknown. Therefore, more components were permitted to be trained for the simulated dataset rather than learning the same number of components as the number of generating sources. The components were set to 14 for rssBSS/sRMSSE and 10 for rKSVD/rACSD/rswsDL, which were applied subject-wise. All algorithms were run for 30 iterations. The ideal initialization for each algorithm was provided after the initialization scenario was assessed using the data, random number generator, and DCT bases separately. For rssBSS, rKSVD, and rACSD, random numbers derived from the standard normal distribution, observed data, and DCT bases were used, respectively.

We experimented with many combinations of the tuning parameters. The ones that yielded the best results regarding the similarity between the ground truth and the retrieved sources were considered. In the instance of rsgICA, twelve components were retained following the second PCA reduction, where the first PCA produced 18 components for each subject. It was discovered that 0.3 and 5000 were the optimal smoothing and sparsity values for rsgICA. The tuning settings for robust ssBSS were $\zeta_1 = 60$, $\lambda_1 = 12$, $K_p = 150$, $g = 0.6$, and $\zeta_2 = 30$. The same settings from rssBSS were applied for sRMSSE, and other parameters were set as $\lambda_2 = 10$, external loop to 10 iterations for subject-specific component extraction, $\zeta_3 = 30$, $\zeta_4 = 6$, $P_m = 7$, $P_c = 7$, and $\lambda_2 = 5$ for common component extraction. The optimal sparsity parameter for rKSVD and rACSD was 8 and 4, respectively. The inner iteration for rKSVD was set to 10 and for rACSD to 5. For rACSD and rswsDL, parameter α was set to 3. In the case of rswsDL, parameter μ was set to 3, and the best sparsity parameters were found to be $\zeta_2 = \zeta_3 = 24$ and $\lambda_2 = 16$.

C. SYNTHETIC DATASET RESULTS

The creation of multi-subject datasets and the learning procedure were conducted several times to show the proposed algorithm's robustness and consistency. In this regard, the experiment was run for i) thirty distinct trials and ii) seven values of ρ from 5 to 17 to progressively increase activation overlaps as seen in Fig. 5a-e.

The source recovery was performed with respect to both group-wise and subject-wise analysis; however, only the results from the subject-wise analysis were plotted, while group analysis results were placed in a table. Following the correlation of each algorithm's trained dictionary atoms/parse code with the ground truth TCs/SMs, underlying source TCs/SMs were obtained by keeping the indices with the highest correlation value. These correlation values, expressed as cTC/cSM, were calculated in reference to ground truth (GT) spatial maps. The mean of the cTC/cSM values over seven spatiotemporal sources were recorded as mcTC/mcSM for each of the seven spatial overlap cases, and their mean over 30 trials and 6 subjects were plotted in Fig. 6 for subject-wise analysis. Fig. 7 displays the convergence rate and the evolution of correlation values for all participating algorithms as function of algorithm iterations. In addition to this, the mean correlation values for group-level extraction of 12 sources are shown in Table 2.

Fig. 6 indicates that for all source recovery situations (spatial/temporal feature and spatial overlap instances), the rswsDL consistently performed better than all other algorithms. Despite high spatial dependence among sources and high temporal noise levels, rswsDL had the best recovery rate. Specifically, although its retrieval capabilities decreased as spatial overlaps grew, it maintained its higher recovery per-

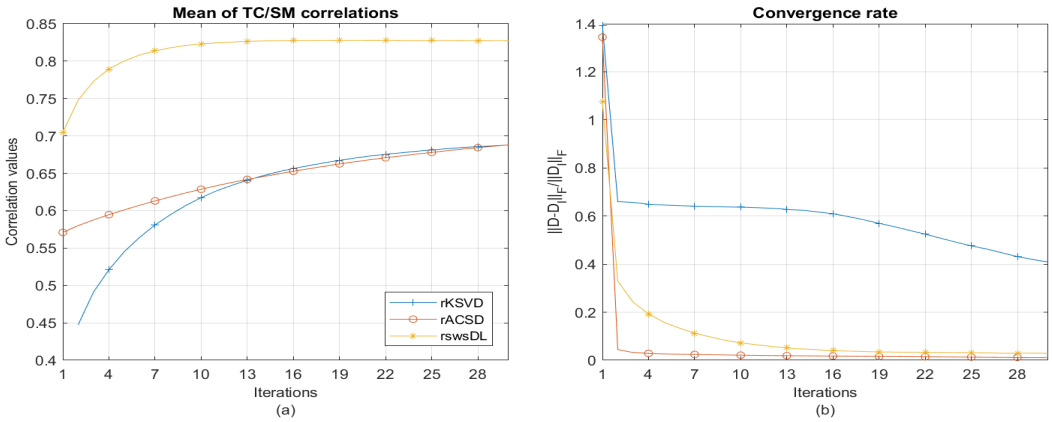

FIGURE 7. For subject-wise dictionary learning, the mean of the a) correlation values and b) convergence rate as a function of algorithm iterations across all subjects.

formance. In comparison, the sRMSSE algorithm emerged as a runner-up but with higher deviations from the mean correlation value for both spatial and temporal recovery cases. Like rswsDL, the rKSVD algorithm exhibited lower standard deviations, but its mean correlation results were not impressive.

From Fig. 7b, it can be inferred that the rswsDL algorithm converged more quickly and required fewer iterations to yield the desired results than rKSVD. This pattern was also seen in Fig. 7a, where the correlation strength for rswsDL almost ceased increasing after the tenth iteration. On the other hand, as the number of iterations rose, both rKSVD and rACSDc continued to show progress in source recovery, whereas the rKSVD method converged slowly.

Table 2 shows the source-wise group analysis results from 30 trials, 7 spatial dependency cases, and 5 algorithms. Overall, it is evident from these results that rswsDL emerged as a winner in source recovery of each spatiotemporal component, with sRMSSE as a runner-up.

D. INTRODUCTION SECTION FIGURE GENERATION

As described in section IV-A, multi-subject synthetic datasets were generated using Laplacian noise. However, different parameter settings were used to generate dataset for section I-A: Laplacian temporal noise variance of 0.6 and spatial noise variance of 0.01, the spread parameter that controls spatial dependence among sources was set to 8, and the lag-1 error minimization term in equation (16) was set to zero for both methods to ensure that autocorrelation maximization does not influence this comparison. The rest of the data generation settings were the same as described in section IV-A.

For an unbiased comparison, tuning parameters, dictionary size, and algorithm iterations were all given the same values for both methods. The selected values for these parameters were set the same as discussed in the section IV-B for rswsDL. For rswsDL$_1$ and rswsDL$_2$, the best sparsity parameters were found to be $\zeta_2 = \zeta_3 = 24$ and $\lambda_2 = 16$, dictionary size of 10 was trained, and algorithm iterations were set to 30 with $\alpha = 1$, and internal iterations set to 4.

E. BLOCK DESIGN DATASET

The 3T MRI raw block design dataset for the motor task was obtained from the HCP's quarter 3 release, whose experimental details are provided in the reference [37], [46], [47]. This dataset, which maps the brain's motor cortex, was acquired in an experiment that lasted 204 secs. The experiment involved giving the volunteers instructions to respond to visual stimuli by tapping their right or left fingers, pinching their right or left toes, or moving their tongues. Subjects completed a 12-second particular movement activity after receiving a three-second visual signal. Ten movement tasks were considered, including two tongue motions, left/right finger, and left/right toe movements. Consequently, there were 13 blocks, of which three were fixation blocks of 15 secs. Six modelled HRFs (MHRs) were created utilizing the canonical HRF and task stimuli linked to five different movement types tongue (T), left toe (LT), right toe (RT), left finger (LF), right finger (RF), and visual type cue (VC) in order to get ground truth TCs.

Each subject's fMRI scan was obtained using a Siemens 3 Tesla (3T) scanner. The following were the acquisition parameters: With 72 contiguous slices and 2 mm isotropic voxels, echo time (TE) = 33.1 ms, TR = 0.72 secs, field of view (FOV) = 208 × 180 mm, flip angle (FA) = 52^o, matrix size = 104 × 90, slice thickness = 2 mm, echo spacing = 0.58 ms, and 284 EPI volumes were gathered where initial 5 scans were discarded. The dataset used in our investigation included eight subjects, ages ranging from 22 to 35.

F. BLOCK DESIGN DATASET PREPROCESSING

As discussed in [37], the SPM-12 toolbox [3] was used for preprocessing the block design dataset. In-depth description of the preprocessing procedures for this dataset, including masking, realignment, normalizing, and spatial smoothing, can be found in [43], [51], and [52]. In order to eliminate motion artifacts, functional images were realigned to the first

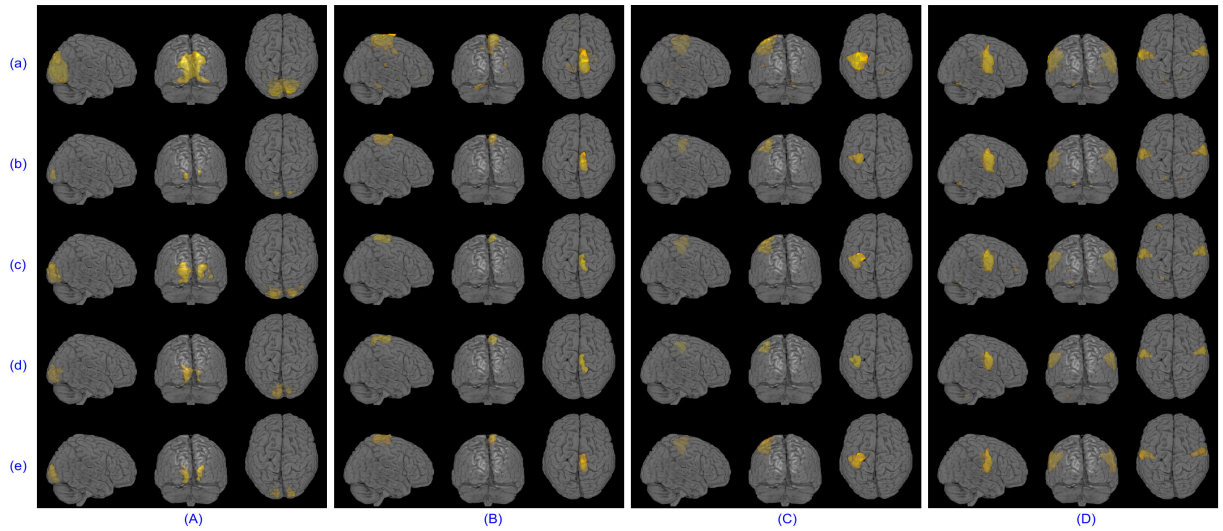

FIGURE 8. For the A) visual cue, B) left toe, C) right finger, and D) tongue tasks in the block design dataset, thresholded common activation maps produced using a) rsgICA, b) sRMSSE, c) rKSVD, d) rACSD, and e) rswsDL at a random field correction $p < 0.001$. Table 3 provides the respective correlation values.

image. Then all images underwent spatial smoothing using a $6\times6\times6$ mm^3 full-width at half-maximum (FWHM) Gaussian kernel after being spatially normalized to a Tailarach template and resampled to $2 \times 2 \times 2$ mm^3 voxels. This was followed by an attempt to remove any data outside the scalp during the masking stage. The four-dimensional dataset was then reformatted and stored as a 2-dimensional matrix termed Y for each m-th individual to be considered a full brain dataset. This step produced dataset Y of size 279×236115 for the m-th subject, where $m = \{1, \ldots, 8\}$. Next, temporal filtering was applied to this Y matrix from all subjects. This included a low-pass filter based on FWHM to remove high-frequency physiological noise and a high-pass filter based on DCT to remove low-frequency trends. An FWHM cutoff of 1 sec was used, and a DCT filter cutoff of 1/150 Hz. Following the previously outlined steps, Y was normalized to have zero mean and unit variance.

G. BLOCK DESIGN DATASET DICTIONARY LEARNING

We tried a wide range of tuning parameter combinations. The ones that produced the best findings in terms of how closely the recovered sources and the ground truth matched each other were taken into consideration. Observed data, random numbers, and DCT bases were used for rKSVD, rssBSS, and rACSD dictionary initialization, respectively. All dictionary learning algorithms ran for a total of 15 iterations, with the exception of the rssBSS and sRMSSE algorithm, which had a total of 30 iterations. After keeping 100 components from the first PCA reduction and 60 from the second PCA reduction, 60 components were retrieved using rsgICA, and its smoothing parameter was set to 5000 while the sparsity parameter to 0.3. In the case of rssBSS, 40 were trained and 60 components from PCA were kept; the remaining parameters were adjusted to $\lambda_1 = 16$, $\zeta_1 = 50$, $K_p = 60$, $g = 0.6$, and $\zeta_2 = 30$. The same rssBSS settings were used for the sRMSSE; the external loop was set to 15 iterations, with $\lambda_2 = 8$ for the extraction of the subject-specific component and $\lambda_2 = 8$ for the extraction of the common component. Its other parameters were set as $K = 100$, $\zeta_3 = 80$, and $\zeta_4 = 8$, $P_m = 30$, and $P_c = 15$. Using rACSD, 40 dictionary atoms were trained for each subject with a sparsity parameter set to 20, inner iterations to 5, and α to 2. The inner iteration was set to 10 and the sparsity parameter to 15 for rKSVD. For rswsDL, 40 atoms were trained, and tuning parameters were set to $\zeta_2 = \zeta_3 = 48$, $\lambda_2 = 25$, $\alpha = 1$, and internal iterations to 5.

H. BLOCK DESIGN DATASET RESULTS

The absence of task-related activation maps in this section led us to select temporal analysis with six constructed MHRs. Similarly, we used Smith's templates because resting state networks do not have temporal profiles. Two methodologies served as the foundation for the analysis: subject-wise and group-level; however, we only discuss group-level analysis for this paper. For the subject-wise analysis, the TCs/SMs for each subject acquired by rKSVD, rACSD, and rswsDL were taken into account, and both common and subject-specific TCs/SMs were considered for sRMSSE. In contrast, the individual TCs/SMs for rsgICA were obtained by back reconstruction. In contrast, the group-level TCs/SMs produced by each competing algorithm served as a reference for group-level assessment. For group-level analysis, the common TCs/SMs for rKSVD, rACSD, and rswsDL were extracted according to the discussions in subsection robust group-level swsDL, they are found in common TC/SM matrices for sRMSSE, and in group-level matrices for rsgICA. These common TCs were eventually correlated with MHRs and averaged SMs with RSNs; the atoms/sparse codes with the most significant correlation values were saved. One can refer to [43] for more details on this extraction.

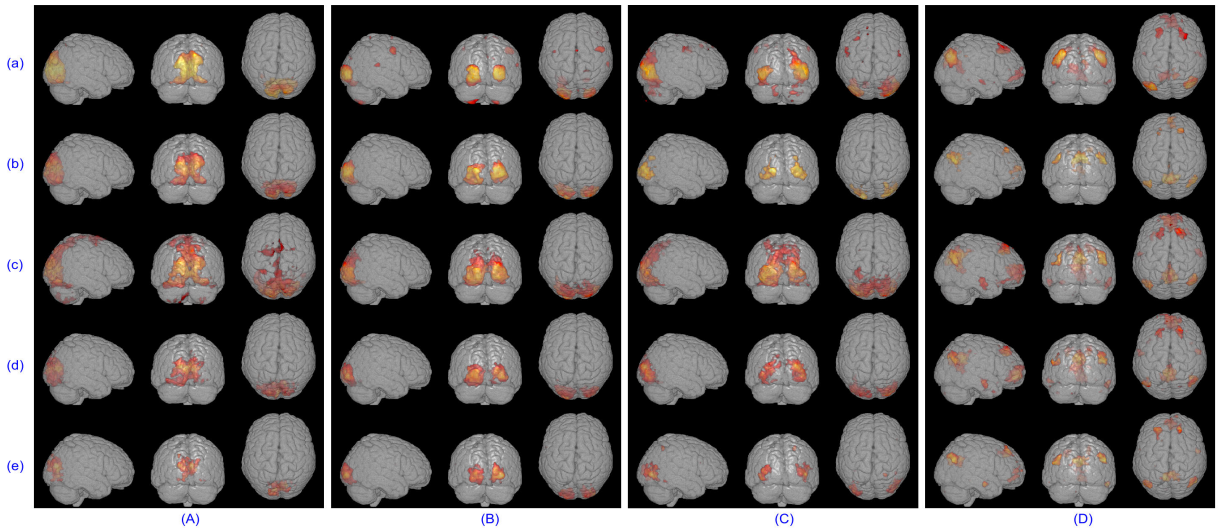

FIGURE 9. Thresholded group-level activation maps derived using a) rsgICA, b) sRMSSE, c) rKSVD, d) rACSD, and e) rswsDL, at a random field correction $p < 0.001$ for the A) medial visual template, B) occipital pole visual template, C) lateral visual template, and D) default mode network template.

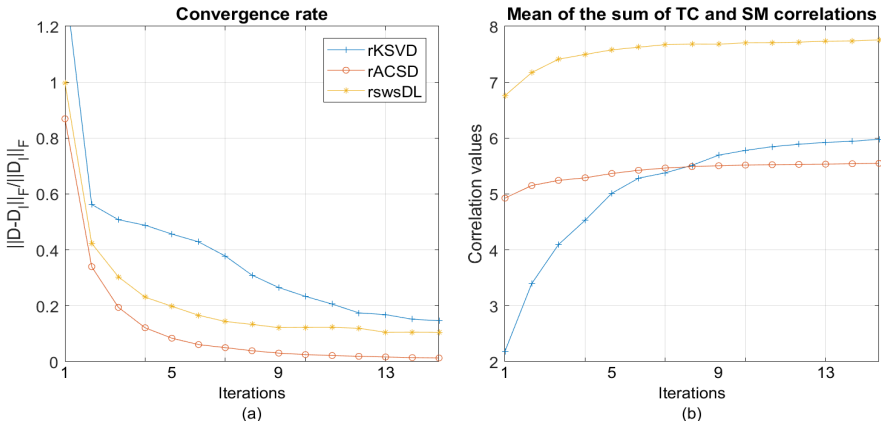

FIGURE 10. The average of the a) correlation values and b) convergence rate as a function of algorithm iterations for three different subject-wise dictionary learning algorithms using block design dataset.

TABLE 3. The highest correlation values between the averaged spatial maps and RSN templates and between group-level dictionary atoms and MHRs are shown for five different algorithms, with the highest values highlighted in bold.

Algos	VC	LT	LF	RT	RF	T	R1	R2	R3	R4	R9	R10	Mean
rsgICA	0.787	0.899	0.866	0.802	0.828	0.882	**0.695**	0.560	0.465	0.394	0.414	0.454	0.670
sRMSSE	**0.937**	0.881	0.836	0.791	0.852	0.842	0.641	0.712	0.487	0.580	0.436	**0.555**	0.713
rKSVD	0.900	0.898	**0.891**	0.800	0.851	0.770	0.516	0.716	0.443	0.575	0.453	0.513	0.694
rACSD	0.891	0.822	0.838	0.671	0.793	0.851	0.624	0.742	0.470	**0.595**	0.420	0.497	0.684
rswsDL	0.926	**0.916**	0.864	**0.854**	**0.879**	**0.890**	0.588	**0.786**	**0.513**	0.555	**0.473**	0.461	**0.725**

For the block design dataset, there were quite a few of activation maps and temporal dynamics, but only a small number have been shown to prevent the lengthening of the study. Fig. 8 displays a sequence of 2D images combined to depict 3D volume for various group-level movement tasks. In the same way, just a portion of the resting-state analysis's findings are shown here. For example, Fig. 9 displays the SMs for the first four resting state network templates. As shown in Table 3, the proposed rswsDL approach performs better than all other algorithms overall, producing atoms and sparse codes with the highest correlation with the ground truth, with sRMSSE coming in second. It is very clear from spatial maps in Fig. 8 that, compared to the other techniques, the activations found by rACSD and rswsDL are more specific to the motor region. Based on Table 3, it can be said that, in terms of correlation values, rswsDL outperformed all other methods. Fig. 9 visually supports this as well, showing that the spatial maps produced by rswsDL are more specific

and consistent in terms of sensitivity than those produced by other algorithms. The correlation strength accumulation and convergence rate of the rKSVD, rACSD, and rswsDL algorithms are shown in terms of iterations in Fig. 10. Here, it can be seen that the rswsDL algorithm progressively outmatches its competitors.

V. CONCLUSION

This paper proposes a novel dictionary learning algorithm to train robust atoms from the dataset containing outliers. This is accomplished by utilizing the loss function that is more effective against the non-Gaussian noise. In place of the Kullback-Leibler divergence, the suggested robust loss function is constructed from the α-divergence. A benefit of the suggested loss is that it is a member of the redescending M-estimator class, which ensures that inferences will remain stable even in the event of significant departures from the Gaussian noise model [34]. Besides, the rotation of reduced dimension to optimal direction avoids the problems associated with model fitting. Another benefit of the suggested algorithm, in contrast to the conventional group analysis, is that it can be used for both task-related and resting-state fMRI data.

The approach used for the suggested algorithm is novel in that it uses the underlying spatiotemporal dynamics from multiple subjects to train dictionary atoms and sparse codes. Because of the spatiotemporal heterogeneity provided by other subjects, this technique enables the incorporation of comparable components from the reduced-dimension space across individuals, leading to higher statistical power. Training the representation matrices and extracting the base/dictionary sparse code solves the suggested model. Synthetic and experimental fMRI datasets have been used to demonstrate the effectiveness of the proposed algorithm, and its performance is consistent and reliable across datasets. The maintenance of the finite basis injective property and a stringent sparsity pattern ensured the convergence of the proposed algorithm [53].

In future, we will attempt to extend the proposed method to hierarchical dictionary learning in order to reduce its computational burden.

ACKNOWLEDGMENT

The block-design dataset was provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

REFERENCES

[1] K. J. Friston, "Modalities, modes, and models in functional neuroimaging," *Science*, vol. 326, no. 5951, pp. 399–403, Oct. 2009.

[2] L. Ma, B. Wang, X. Chen, and J. Xiong, "Detecting functional connectivity in the resting brain: A comparison between ICA and CCA," *Magn. Reson. Imag.*, vol. 25, no. 1, pp. 47–56, Jan. 2007.

[3] W. Penny, K. Friston, J. Ashburner, S. Kiebel, and T. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. New York, NY, USA: Academic, 2007.

[4] A.-K. Seghouane and M. U. Khalid, "Hierarchical sparse brain network estimation," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Sep. 2012, pp. 1–6.

[5] M. U. Khalid and A.-K. Seghouane, "A single SVD sparse dictionary learning algorithm for FMRI data analysis," in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, Jun. 2014, pp. 65–68.

[6] M. U. Khalid and A.-K. Seghouane, "Constrained maximum likelihood based efficient dictionary learning for FMRI analysis," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 45–48.

[7] A. Shah, M. U. Khalid, and A.-K. Seghouane, "Recovering HRFs from overlapping Rois in FMRI data using thresholding correlations for sparse dictionary learning," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 5756–5759.

[8] M. U. Khalid and A.-K. Seghouane, "Unsupervised detrending technique using sparse dictionary learning for FMRI preprocessing and analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 917–921.

[9] A.-K. Seghouane and M. U. Khalid, "Learning dictionaries from correlated data: Application to FMRI data analysis," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2340–2344.

[10] G. K. Aguirre, E. Zarahn, and M. D'Esposito, "The variability of human, bold hemodynamic responses," *NeuroImage*, vol. 8, no. 4, pp. 360–369, Nov. 1998.

[11] M. J. McKeown, T.-P. Jung, S. Makeig, G. Brown, S. S. Kindermann, T.-W. Lee, and T. J. Sejnowski, "Spatially independent activity patterns in functional MRI data during the stroop color-naming task," *Proc. Nat. Acad. Sci. India A, Phys. Sci.*, vol. 95, no. 3, pp. 803–810, Feb. 1998.

[12] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Exploratory FMRI analysis by autocorrelation maximization," *NeuroImage*, vol. 16, no. 2, pp. 454–464, Jun. 2002.

[13] M. U. Khalid, A. Shah, and A.-K. Seghouane, "Adaptive 2DCCA based approach for improving spatial specificity of activation detection in functional MRI," in *Proc. Int. Conf. Digit. Image Comput. Techn. Appl. (DICTA)*, Dec. 2012, pp. 1–6.

[14] M. U. Khalid and A.-K. Seghouane, "Improving functional connectivity detection in FMRI by combining sparse dictionary learning and canonical correlation analysis," in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 286–289.

[15] A. Shah, M. U. Khalid, and A.-K. Seghouane, "Comparing causality measures of FMRI data using PCA, CCA and vector autoregressive modelling," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 6184–6187.

[16] M. U. Khalid, A. Shah, and A.-K. Seghouane, "Sparse dictionary learning for FMRI analysis using autocorrelation maximization," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 4286–4289.

[17] W. Lin, H. Wu, Y. Liu, D. Lv, and L. Yang, "A CCA and ICA-based mixture model for identifying major depression disorder," *IEEE Trans. Med. Imag.*, vol. 36, no. 3, pp. 745–756, Mar. 2017.

[18] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, Jun. 2000.

[19] O. Friman, J. Cedefamn, P. Lundberg, M. Borga, and H. Knutsson, "Detection of neural activity in functional MRI using canonical correlation analysis," *Magn. Reson. Med.*, vol. 45, no. 2, pp. 323–330, Feb. 2001.

[20] A. Krishnan, L. J. Williams, A. R. McIntosh, and H. Abdi, "Partial least squares (PLS) methods for neuroimaging: A tutorial and review," *NeuroImage*, vol. 56, no. 2, pp. 455–475, May 2011.

[21] J. Särelä and R. Vigário, "Overlearning in marginal distribution-based ICA: Analysis and solutions," *J. Mach. Learn. Res.*, vol. 4, pp. 1447–1469, Dec. 2003.

[22] M. Li, Y. Liu, F. Chen, and D. Hu, "Including signal intensity increases the performance of blind source separation on brain imaging data," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 551–563, Feb. 2015.

[23] Y. Zhang, M. Li, H. Shen, L.-L. Zeng, and D. Hu, "A robust multi-subject FMRI analysis method using dimensional optimization," *IEEE Access*, vol. 7, pp. 125762–125770, 2019.

[24] J. Liu, X.-C. Tai, H. Huang, and Z. Huan, "A weighted dictionary learning model for denoising images corrupted by mixed noise," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1108–1120, Mar. 2013.

[25] P. J. Huber, "Robust statistics," in *International Encyclopedia of Statistical Science*. Berlin, Germany: Springer, 2011, pp. 1248–1251.
[26] C. Lu, J. Shi, and J. Jia, "Online robust dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 415–422.
[27] W. Jiang, F. Nie, and H. Huang, "Robust dictionary learning with capped l1-norm," in *Proc. 24th Int. Conf. Artif. Intell.*, Jul. 2015, pp. 3590–3596.
[28] S. You, C. Xu, and C. Xu, "Online dictionary learning with confidence," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 707–716.
[29] V. Fritsch, G. Varoquaux, B. Thyreau, J.-B. Poline, and B. Thirion, "Detecting outliers in high-dimensional neuroimaging datasets with robust covariance estimators," *Med. Image Anal.*, vol. 16, no. 7, pp. 1359–1370, Oct. 2012.
[30] V. Fritsch, G. Varoquaux, J.-B. Poline, and B. Thirion, "Non-parametric density modeling and outlier-detection in medical imaging datasets," in *Machine Learning in Medical Imaging*, F. Wang, D. Shen, P. Yan, and K. Suzuki, Eds. Cham, Switzerland: Springer, 2012, pp. 210–217.
[31] A. F. Mejia, M. B. Nebel, A. Eloyan, B. Caffo, and M. A. Lindquist, "PCA leverage: Outlier detection for high-dimensional functional magnetic resonance imaging data," *Biostatistics*, vol. 18, no. 3, pp. 521–536, Jul. 2017.
[32] Y. Alemán-Gómez, A. Arribas-Gil, M. Desco, A. Elías, and J. Romo, "Depthgram: Visualizing outliers in high-dimensional functional data with application to FMRI data exploration," *Statist. Med.*, vol. 41, no. 11, pp. 2005–2024, May 2022.
[33] S. Mukherjee, R. Basu, and C. S. Seelamantula, "L1-K-SVD: A robust dictionary learning algorithm with simultaneous update," *Signal Process.*, vol. 123, pp. 42–52, Jun. 2016.
[34] A. Iqbal and A.-K. Seghouane, "Robust dictionary learning using a-divergence," in *Proc. ICASSP - IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2972–2976.
[35] D. Sahoo and C. Davatzikos, "Learning robust hierarchical patterns of human brain across many FMRI studies," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 29034–29048.
[36] J. Huang, E. Busch, T. Wallenstein, M. Gerasimiuk, A. Benz, G. Lajoie, G. Wolf, N. Turk-Browne, and S. Krishnaswamy, "Learning shared neural manifolds from multi-subject FMRI data," in *Proc. IEEE 32nd Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Aug. 2022, pp. 01–06.
[37] M. U. Khalid and M. M. Nauman, "A novel subject-wise dictionary learning approach using multi-subject FMRI spatial and temporal components," *Sci. Rep.*, vol. 13, no. 1, p. 20201, Nov. 2023.
[38] A.-K. Seghouane and A. Iqbal, "Consistent adaptive sequential dictionary learning," *Signal Process.*, vol. 153, pp. 300–310, Dec. 2018.
[39] M. U. Khalid, B. A. Khawaja, and M. M. Nauman, "Efficient blind source separation method for FMRI using autoencoder and spatiotemporal sparsity constraints," *IEEE Access*, vol. 11, pp. 50364–50381, 2023.
[40] S.-I. Amari, "Differential geometry of statistical models," in *Differential-Geometrical Methods in Statistics*. Cham, Switzerland: Springer, 1985, pp. 11–65.
[41] M. Aharon, M. Elad, and A. Bruckstein, "*rmK*-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[42] A. Iqbal, A.-K. Seghouane, and T. Adali, "Shared and subject-specific dictionary learning (ShSSDL) algorithm for multisubject FMRI data analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 11, pp. 2519–2528, Nov. 2018.
[43] M. U. Khalid, "Sparse group bases for multisubject FMRI data," *IEEE Access*, vol. 10, pp. 83379–83397, 2022.
[44] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, and C. F. Beckmann, "Correspondence of the brain's functional architecture during activation and rest," *Proc. Nat. Acad. Sci. India A, Phys. Sci.*, vol. 106, no. 31, pp. 13040–13045, Aug. 2009.
[45] E. B. Erhardt, E. A. Allen, Y. Wei, T. Eichele, and V. D. Calhoun, "SimTB, A simulation toolbox for FMRI data under a model of spatiotemporal separability," *NeuroImage*, vol. 59, no. 4, pp. 4160–4167, Feb. 2012.
[46] D. C. Van Essen et al., "The human connectome project: A data acquisition perspective," *NeuroImage*, vol. 62, no. 4, pp. 2222–2231, Oct. 2012.
[47] D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, D. Nolan, E. Bryant, T. Hartley, O. Footer, J. M. Bjork, R. Poldrack, S. Smith, H. Johansen-Berg, A. Z. Snyder, and D. C. Van Essen, "Function in the human connectome: Task-FMRI and individual differences in behavior," *NeuroImage*, vol. 80, pp. 169–189, Oct. 2013.
[48] V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Hum. Brain Mapping*, vol. 14, no. 3, pp. 140–151, Nov. 2001.
[49] Z. Boukouvalas, Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, "Sparsity and independence: Balancing two objectives in optimization for source separation with application to FMRI analysis," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1873–1887, Mar. 2018.
[50] M. U. Khalid, M. M. Nauman, M. I. Petra, and A. Iqbal, "Improving source separation for multi-subject FMRI data by incorporating signal intensity and spatiotemporal basis expansion," in *Proc. Manuscript Accepted Publication Int. Conf. Image Graph. Process.*, 2024.
[51] C.-M. Ting, A.-K. Seghouane, M. U. Khalid, and S.-H. Salleh, "Is first-order vector autoregressive model optimal for FMRI data?" *Neural Comput.*, vol. 27, no. 9, pp. 1857–1871, Sep. 2015.
[52] M. U. Khalid, "Dictionary learning algorithms for functional magnetic resonance imaging," Ph.D. dissertation, College Eng. Comput. Sci., Austral. Nat. Univ., Canberra, ACT, Australia, 2015.
[53] K. Bredies and D. A. Lorenz, "Linear convergence of iterative soft-thresholding," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 813–837, Dec. 2008.

MUHAMMAD USMAN KHALID received the M.Sc. degree in electrical engineering from the Royal Institute of Technology, Stockholm, in 2010, and the Ph.D. degree in engineering and computer science from Australian National University, Canberra, in 2015. He was a Visitor Academics with the Department of Electrical and Electronic Engineering, The University of Melbourne, and a Research Fellow with the School of Mathematics and Statistics, The University of Melbourne. Currently, he is an Assistant Professor with the Department of Information Systems, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. His research interests include statistical signal processing, numerical optimization, blind source separation, and medical imaging.

BADER M. ALBAHLAL received the bachelor's degree in information systems from Imam Mohammed Ibn Saud Islamic University and the Ph.D. degree in information from Florida State University.

He boasts over five years of valuable experience in the private and public sectors. He is an accomplished Assistant Professor with a background spanning both private industry and academia. His research is dedicated to exploring the intricate dynamics of technology adoption and its profound influence on stakeholders. His scholarly pursuits delve into the realms of data science, cybersecurity, and the adoption of emerging technologies.